



# PERSONAL LIBRARY SOFTWARE



Because PL is so powerful and flexible, you can design an information product that anyone can use successfully. The first-time or occasional user will be able to sit down and find what they need with no training.

PL gives you the flexibility to design your product exactly the way you want it. Choose from character-based or state-of-the-art graphical interfaces. Use our standard products... or build your own interface using PLS's powerful search engine, CPL.

## Taking Advantage of PL's Power You Get

*Ease of Use Through Smart Searching:* PL's unique smart technology and flexible interface design tools make finding even the most obscure information a snap. PL liberates your users from unnatural query language.

*MS-Windows* or windows-like character-based interfaces propel your product into the Nineties with state-of-the-art look-and-feel.

**Print Media Analogy:** Your PL publication can have the look and feel of an electronic book. With PL, your readers can open a Table of Contents, point and click to a particular section or page and browse page by page.

**Fill in the Blanks Forms:** Your subscribers can complete complex searches by reading easy-to-follow instructions and typing in their queries the way they would fill in a form.

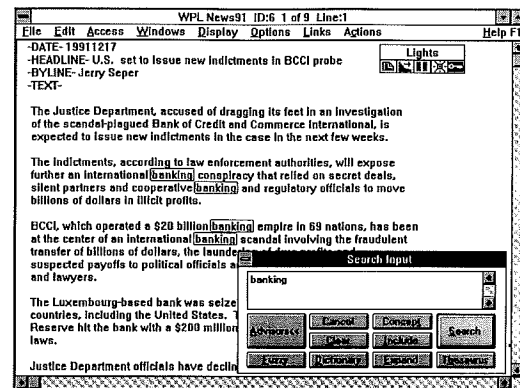
**Field Organization:** Up to 255 variable-length fields per record permit you to bring a full-text approach, PL Relevancy Ranking Power, and structured data precision together into a single information base.

**Embedded Hypertext:** Giving your users a way to navigate through associations, PL allows you to make logical links to other records, graphics or even other executable programs.

**Point and Shoot Searching:** You prepare a window of information that can be browsed with industry-standard commands. Your reader finds a subject (from a single word to a phrase), points and double-clicks and either locates all of the relevant documents or jumps to a precise place in the text.

**Fully Customizable Interface:** PL puts all of its multiple functionalities in your designers' hands. CPL, an object oriented Application Programmer's Interface, provides your unique product with the smartest technology available today.

**Full Text Searching** in PL requires no contorted Boolean Logic searching, no more confusion among AND, ORs and NOTs. Natural language queries and PL's powerful Probable Relevance Ranking help the user to locate the most relevant information quickly.



## Personal Librarian Search Screen

## Provide Your Users With Intelligent Retrieval And Analysis Tools

- Relevance Ranking
- Root matching
- Search-by-example
- Spontaneous concept search
- Automatic word associations
- Thesauri
- On-line help
- Natural language queries

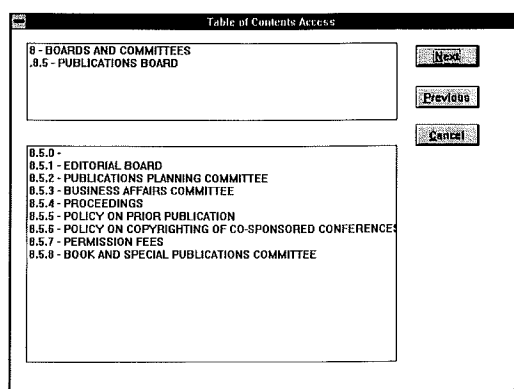
*Integrate With:*

- desktop publishing
- Graphics/image processing, sound
- CD-ROM, optical, magnetic media, jukeboxes
- Macintosh, Windows, DOS, UNIX, VMS, networks
- Multiple/distributed databases

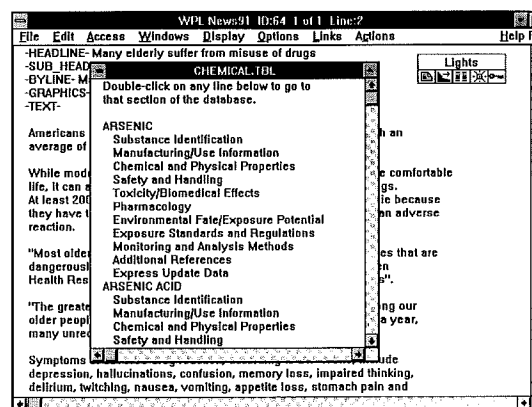
[illegible]

The following subjects are covered in PLS's database design and build documentation:

- Database Sources
- Database Design
  - *Fields*: none, purpose, types, multiple values, repeating fields, field-groups, interrecord relationships, multiple record types, search/display/list attributes
  - *Document Chunking*: What is a document? strategies: page, chapter, whole items
  - *Distributed Architectures*: spreading out source files, LANs, integrating static and dynamic databases
  - *Data Preparation*: tools, procedures
- Preparation for Different Search Modes
  - Windows, positions, sizes, fonts
  - Forms for query, display, reports
  - Guided (Index) and general view files: search aids
  - Hypertext Links, naming/storing
- Hierarchies for Special Title Lists, Tables of Contents, and Domains
- Preparation for User Assistance
  - Thesauri
  - Customizable Help
  - Start-up screens
- Hypertext Links
  - Image capabilities/techniques
  - Point-to-field strategies
  - Command-link capabilities/strategies
  - Search-links
- Encryption and Access Control
- Configuration
  - Environment variables, installation, memory
  - Run-time parameters
  - Initialization: modes, parameters, defaults



Personal Librarian Table of Contents Screen



Personal Librarian Index Screen

- Indexing
  - Stop word list, stemming
  - Definition file
  - Placement of files in alternative locations
  - PL Admin/separate utilities
  - Big build strategies
  - Validate
  - Verify
- CD-ROM Considerations
  - Layout installation, caching, tuning

\* The screen shots shown here are standard WPL screens.

Remember, PL excels at what a searcher needs most: finding information. PL makes your information more useful and more valuable. What is it worth to your users to find information in seconds that would otherwise have taken days, or might never have been found? PL is a very fast way to search and retrieve information; it is by far the fastest way to bring users and *relevant* information together.

Give your clients PL with your information -- they'll find out what they're not missing.



2400 Research Boulevard, Suite 350  
Rockville, MD 20850-3243  
301/990-1155

Personal Librarian and Personal Librarian for Windows are tradenames of Personal Library Software. All product names are trademarks of their respective manufacturers.



## The Most Effective Retrieval Engine Available

- Powerful And Fast Retrieval
- Highly Portable
- Designed For Fast Application Development
- Proven

CPL (Callable Personal Librarian) is a C language application programmer's interface (API) for speeding development of sophisticated full-text database applications or integrating full-text retrieval into your existing application (image management, DBMS, etc.).

It is the same API used by Personal Library Software's PL and WPL end-user retrieval applications. CPL includes both the document indexing functions and the search and retrieval functions.

CPL is the most effective full-text retrieval engine on the market, incorporating proven relevance ranking and association heuristics that let the end user effectively search vast information resources.

CPL is written entirely in "C" and "C++" and is designed to run under UNIX, VMS, IBM-compatible microcomputers under DOS and MS Windows, and on the Apple Macintosh. CPL has no special library requirements other than the host system's C run-time library (CRTL).

CPL's object-based architecture offers conceptual simplicity to the application programmer. The "stateful" features ease the workload when creating direct-manipulation, graphical user interfaces.

CPL has proven reliability; it has been in production use since 1987 on major platforms (VMS, UNIX, DOS, MS Windows, Mac OS).

## The Powerful Query Language Provides

**Relevance ranking:** Documents retrieved by a query are ordered by computed probability of relevance to the query. This is a proven, powerful technique that provides significant leverage to the user for dealing with large, complex informa-

tion resources. This technique is the cotton gin of full-text management!

**Standard Boolean operators:** The traditional way to search text with the AND, OR, and NOT connectors is available at any time.

**Proximity/adjacency:** Both proximity (within N words) and adjacency are available for precision searching.

**Statistical thesaurus:** One of the major problems in text searching is finding the "right" words to use in a query. CPL has a set of tools that suggest possible new words to use based on words or documents the user already knows. These suggested words offer new pathways to additional information in the database. There is no preprocessing or specialized indexing required to use this feature!

**Document as a query:** Very often the searcher can find one useful document, but has trouble finding more. CPL has a query operator that means "find me other documents like this one." A real-time, on-the-fly analysis is done to locate similar documents.

**Field searches:** Documents can be structured into fields. Whole searches and search terms (explicitly or by default) can be limited to named fields.

**Field displays:** Each of several display modes can be limited to named fields, by defaults, or interactively.

**Wildcard search terms:** Multiple "don't care" and single "don't care" query characters allow searching for word variants.

**Word stemming:** CPL will automatically include variations of single or multiple word bases in the query to increase the likelihood of finding relevant documents (e.g., singular, plural and verb endings).

**Synonym thesaurus:** CPL supports the use of a user/administrator defined synonym thesaurus. A query operator tells CPL to "replace this word with its synonyms from the thesaurus."

**Past queries:** Reuse past queries in combination or to form new queries to narrow or broaden a search.

**Range search:** For numeric quantities, complete range searching operations are allowed.

## Other Features

**Foreign language support (international character set):** All language dependencies are removed from CPL into configuration files (resources). CPL is currently being used in French, Italian, Spanish, and German language applications.

**Fast indexing and updating:** On the Sun 4, CPL indexes about 50 megabytes of source text per hour. On a 33 MHz 386 PC, 8 megabytes per hour is common with extra memory and disk caching. For updating, CPL only modifies those portions of the index that need it, so updating can keep pace with a very dynamic database.

**Distributed databases:** CPL allows distribution of text files and index files across multiple devices and media (CD-ROM, WORM, all standard magnetic, LANS).

**CD-ROM performance:** CPL is optimized for fast retrieval on CD-ROM as well as on magnetic media.

**Hypertext links:** CPL directly supports typed hypertext links.

**Virtual databases:** An application can view all open databases as a single database.

**Configurability:** Your application has control over space/time trade-offs. CPL offers run-time performance tuning and application control of memory usage.

**Flexible administration:** Database indexing and administration can be handled by the same utility programs shipped with WPL and PL.

**Distributed, multi-user update with transaction integrity and restart/recovery:** Concurrency control protects retrieving users from modifications to the database until they are complete. Transaction integrity ensures that an abort during an index update will not affect the database. These capabilities are coupled with sophisticated storage management that minimizes index overhead and optimizes the update and retrieval speeds.

**Document classes:** If you have a document architecture that is currently not directly supported by PLS, you can write a small piece of document management code and plug it into CPL. CPL can then index and retrieve the document in its native format so no conversion is required!

## Availability

- VMS 5.3 and later (uses QIO for high-speed IO)
- MS DOS 3.3 and later
- MS-Windows 3.x as a DLL
- Sun 4 Sun OS 4.0
- Macintosh
- Many other UNIX platforms

Available as a retrieval-only run-time library, or with full indexing capability (record add, delete, update).

## System Hardware Requirements

- IBM PC 386 or higher  
2MB memory or more  
Hard Disk  
MD DOS 3.3 or later
- Sun 4
- VAX
- Mac II and up

## System Specifications

- Maximum number of records: 16 million
- Maximum fields per record: 255
- Maximum record size: 2Gb
- Maximum database size: main index limited to 2Gb
- Index overhead: ~65%



2400 Research Boulevard, Suite 350  
Rockville, MD 20850-3243  
301/990-1155



Find What You're Missing

PERSONAL LIBRARY SOFTWARE



## The Quest For A "Smart" Approach To Text Searching

Today's fast and relatively inexpensive desktop computers, equipped with large data storage devices like optical disks often networked into vast data sources, *can* put unprecedented amounts of information at the user's fingertips. What is needed is a tool that is both simple to use and powerful enough to help turn data into knowledge.

Historically, even the most experienced information searchers of large text databases have been frustrated by the lack of adequate retrieval tools. Too often, "You don't know what you're missing," has proven to be more than just a slogan. Academic studies of the effectiveness of Boolean keyword systems have consistently shown that these traditional systems miss an average of 80% of relevant materials.

There are some simple explanations of the essential problems. One of the major culprits is something that we value most highly, that is, the rich and subtle nature of the way we use language. For years, there has been a quest for a "smart" approach to text searching that takes into account the conceptual nature of information and provides tools to accommodate this diversity.

## Other Approaches Have Been Tried -- With Limited Success

Some "expert system" or rule-based approaches to this problem have been attempted but have had extremely limited impact because of their inherent rigidity and difficulty to set up. They, like the traditional (Boolean only) search engines, tend to work best in static databases with carefully limited vocabulary sets. But even here, they tend to be blind to the concepts and trends that can only emerge in the process of understanding a body of knowledge.

In today's fast moving, speedily changing world, decision makers in business and government also need to cope with massive amounts of constantly shifting information flows. By the time information searchers have been assigned to an information retrieval expert, it may be too late! What is worse, the person who is most intimately involved in the body of knowledge with which the needed information is associated -- and who is most able to understand what is relevant-- is kept at arm's length from the search process.

Clearly, a tool that combines a simple, "smart" approach to searching, the brain power of the searcher, and the high speed of today's desktop computers becomes invaluable! And, in fact, the tool does exist: PL uses computing power in a very different way that is simple, elegant and, above all, effective! Through its unique heuristic algorithms *that require no preprocessing on the part of the information base organizer*, PL is able to locate the missing information necessary to make truly informed judgements and decisions.

And this smart searching requires no extra work on the part of the user. PL makes searching easier than traditional or knowledge-based systems. PL also has a full array of precise phrase searching tools. The searcher has at his or her fingertips -- or at the click of a mouse -- Boolean, proximity, field and refinement operators, search-by-example, and concept searching tools that can be used separately or in concert.

## How PL Tracks Down "Nicaraguan Contras"

To understand how PL Concept searching works, it is necessary to see it in action. For that purpose, let's follow a search being carried out in a sample database made up of a collection of newspaper articles provided by one of PL's customers. The database contains several thousand articles that appeared over a period of weeks in 1988.

Our searcher is a reporter who is researching the status of events in Nicaragua over a certain crucial period of time. She chooses "Nicaraguan Contras" as her search terms and gets back nine articles. The documents containing both key terms are ranked higher than those containing only one of the terms. PL automatically employs its stemming algorithm for her search and includes hits of "Nicaragua" and "Contra" as well. A list of the headlines of the articles she finds looks like this:

- 1 Contra Aid Bill Passes House in 345-70 Vote
- 2 Sandinistas Free 100 Prisoners
- 3 Reagan Finally Gets Contra Relief Bill
- 4 Wright Aide Tried His Own Arms Deal
- 5 Contra Disarmament on Hold
- 6 Calderon: Arias Soft on Communism
- 7 Remember Who Made the Sandinistas Winners
- 8 Cease-Fire Likely to Wound Contras
- 9 Nicaraguan Exiles Cry 'Betrayal'

After browsing through these documents, our reporter decides to look for *conceptually* related information by using the Expand function. After clicking on Advisors in the Search window, she types the keywords "Nicaragua and Contras" and clicks on Expand.

In a matter of seconds, a window on the screen announces that PL has located 64 "Expanded" words. The algorithm has determined this list on the fly based upon information in the articles themselves - *no preparatory work at all is involved*.

The following terms appear in the Expand window:

Humanitarian	Jennifer
Calero	Foley
Cheney	Marxist
Contra	Contra
Calderon	Indictments
Package	Iran
Lethal	North
Sandinista	
Ollie	

By selecting such suggested key words as "indictments" or "Ollie," with the click of the mouse, a newly "redefined" search is generated. Our enterprising reporter is able to explore related, relevant information not present in any of the articles that contained the words "Nicaragua and Contras."

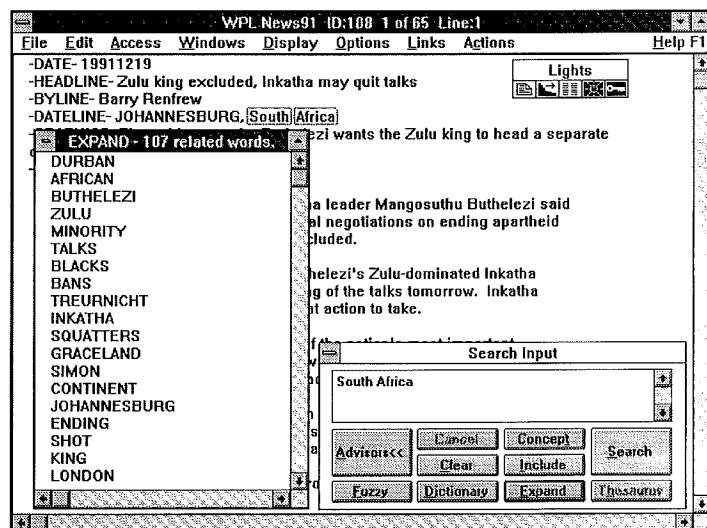
In addition, the searcher can decide to use a more automatic version of the Concept Search capability. In this case, the searcher types the key words into the standard search box, and instead of pressing the search key, clicks on the concept button. The concept feature has been specially designed to take advantage of the type of automatically generated word lists produced by the Expand function. With a concept search, a group of these automatically selected words is combined with the original key words in a full-text search and then is processed through PL's Relevancy Ranking algorithm. In this case, articles about Oliver North, John Poindexter, the Sandinistas, humanitarian aid, and the Contras are all made available for easy browsing with *no preprocessing*.

In another complementary approach to concept searching (sometimes called relevance feedback), the searcher finds an interesting record and types the ID number assigned by PL into the Search Box. With this type of information, PL automatically assumes that the user is looking for conceptually similar information and takes a simple search command to mean it should apply its smart technology. The searcher can use whole records, cut-and-pasted passages, "tagged" words, "Expand" and "Concept" (and all of the traditional operators) in full combination -- whatever helps get the job done.

### Find Out What You're Missing

With tools this powerful, the flexibility of a truly refined system, and the interactivity of a specially designed Windows tagging feature that allows suggested keywords to be applied to a search at the clicking of the mouse, PL finds what other less powerful tools would miss.

It's not surprising that we say PL is the only smart way to find what you are looking for.



Personal Librarian Expand Window



2400 Research Boulevard, Suite 350  
Rockville, MD 20850-3243  
301/990-1155



Through a broad array of innovative and yet proven technologies, Personal Librarian offers the most advanced information management system available today. Personal Librarian (PL) is particularly well-designed for the management of large integrated collections of textual and graphic information. *And with all its power, flexibility, and speed, even an inexperienced researcher can use it effectively the first time -- every time.*

## What Can PL Do For You?

PL has been designed to meet the growing need for Document Storage and Retrieval, made feasible today through the integration of optical storage media, scanning and optical character recognition, powerful but inexpensive desktop computers efficiently networked, and electronic publishing media like CD-ROM. PL technology is currently being used by:

- Law Firms for litigation support and practice area knowledge systems
- Newspapers for electronic archives
- Corporate marketing departments to track competitors
- Manufacturers and engineers for their technical documentation
- Pharmaceutical firms to support approval testing
- Law enforcement agencies to find vital missing clues
- Major Defense and Civil Government Agencies
- International On-line Information Systems for publication of government information
- Public utilities to manage regulatory correspondence
- Publishers of scholarly journals for CD-ROM distribution
- Universities for researching full-text books and manuscripts
- Libraries for full-text books, multi-media manuscripts, periodicals, and catalogs

Every day people are finding new uses for this powerful information manager within the following broad categories:

**Electronic Document Storage and Retrieval.** File cabinets, even boxcars full of your most vital documents, can be efficiently searched with instant, precise retrieval using standard off-the-shelf hardware and software.

**Electronic Publishing.** Volumes of information can be distributed over networks or on optical disks such as CD-ROM. PL's powerful array of electronic publishing, authoring, and retrieval tools are the critical link in making a large information base into a useful product.

**Automated Message Handling and Retrieval.** PL's fast indexing and updating functions can keep up with a rapid influx of new information and quickly make it available to

those people in your organization who most need to know. Users can be logged onto the network for simultaneous searching of information as the database is updated.

## Practical And Efficient Applications

PLS is distributed in nine countries by a growing network of specialized, value-added resellers who often integrate their own know-how and technologies with PL. In addition to standard off-the-shelf PL support, PL resellers provide:

- Fully-integrated document management systems that also integrate scanning, optical character recognition (OCR), networking, and optical storage technologies
- CD-ROM database design, preparation, and pre-mastering services
- Legal Information Retrieval Systems for both litigation support and internal law firm document, work product, and information management
- Integration services for inclusion of PL's superior search capability into vertical applications for newspapers and libraries

## The Core Technology

From the beginning, PL was designed to be flexible. Unlike other retrieval systems that are suitable for only one type of application (for example, technical manuals) or one medium (like CD-ROM), PL adapts to many varied environments and applications. The casual user will find that PL is very easy to use and requires little additional learning. Simply say what you want and then point and click. The experienced application designer or database administrator will find that PL is a powerful core technology and set of tools that function in many environments.

## PL's Distinctive Features Include

- Intelligent search and relevance-ranked retrieval
- Heuristic, expert search assistance
- Simple, yet powerful user interface
- Full functionality
  - Text, data, and image
  - Hypertext
  - Access via fill-in-the-blank forms, table of contents, keyword, Boolean, or concept search
  - Printing, sorting, access control, session logging, and broad utility functions
- Callable version for embedding in other applications or for custom interfaces

- *Portable* across many computers and operating systems (DOS, MS-Windows, UNIX, VMS, MAC)
- *Speed*: fast access whether on CD-ROM, WORM or magnetic disk, PC, or super workstation. Tunable run-time parameters let you optimize for your environment (e.g., storage media, available memory).

## Concept + Interaction = The Smarter Searcher

At its core, PL is the most powerful engine for finding relevant information on the market today. PL achieves superior recall and precision by using efficient algorithms that have successfully adapted heuristic principles to today's computing environment. These functions go to the heart of the problem of finding information in large text databases *even when you don't know the "right" words to use*. Traditional Boolean searchers find themselves trying to "narrow" their search by using exclusionary word and logic formulations. If they don't, their hit lists are too large. Unfortunately, this "narrowing" approach is a major reason why searchers often miss 80% of the relevant information.

Instead of narrowing, PL *focuses*. PL's unique Probable Relevance Ranking *automatically* orders information according to the terms you used in your search. The top-ranked items are most likely to be relevant. Combined with PL's hit list summary, the time it takes to find relevant information is greatly reduced.

PL's unique Expand function can examine a huge database and in seconds deliver a list of terms that help identify relevant information. Sometimes words of no immediately apparent connection can turn up the most vital information.

PL can find vitally important information about a subject even when the key words you use for your search are not present in the retrieved document. It uses heuristics to locate related words, which it then uses for a comprehensive search, PL then re-analyzes the results and presents them in an easily browsable format with the highest ranked documents at the top of the list.

## System Requirements

**WPL requires a PC that will run Microsoft Windows.**  
**Index overhead is about 65% for archived databases.**  
**Overhead is higher for very dynamic databases.**

- Maximum records per database: 16 million
- Maximum fields per record: 255
- Maximum record size: 2Gb
- Maximum database size: 2Gb index

## Product Specifications

PL is available on PC (DOS and MS-Windows), Macintosh, VAX, Sun and other UNIX platforms.

### Features:

Smart relevance ranking  
 Multiple database searching with merged results  
 Linear and Hierarchical browsing  
 Domains  
 Hypertext  
 Images (CCITT 3/4)  
 Cut-and-paste  
 Link to other applications  
 Full-text / field specific searching  
 Boolean, proximity, wild-card operators  
 Numeric range searching  
 Root and/or exact word match  
 Activity trail  
 On-line help (customizable)  
 Print options  
 Statistical (automatic) and manual thesauri  
 Search by documents, passages, tagged words, and phrases  
 Foreign languages: supports international character set, interfaces available in French and Italian (Inquire about Spanish and German)  
 Fast, dynamic database updating  
 Access control  
 Distributed databases - multiple files, drives, devices; network compatible  
 Run-time configurable to tune performance for computing environment  
 Configurable interfaces  
 Full API

Not all features are available on all platforms.



**2400 Research Boulevard, Suite 350**  
**Rockville, MD 20850-3243**  
**301/990-1155**